Hen's Egg Test – Chorioallantoic Membrane Test Method

[This Page Intentionally Left Blank]

IV. HEN'S EGG TEST-CHORIOALLANTOIC MEMBRANE TEST METHOD

1.0 HET-CAM TEST METHOD RATIONALE

1.1 Scientific Basis for the HET-CAM Test Method

1.1.1 Mechanistic Basis of the HET-CAM Test Method

The rabbit eye is the current reference standard in predicting what will happen when the human eye is directly exposed to a chemical, even though the rabbit eye is somewhat structurally different from the human eye. It should always be noted, however, that suitable human data would be vastly preferred as a comparative standard. The chorioallantoic membrane (CAM) contains vascular membrane structures. The Hen's Egg Test – Chorioallantoic Membrane (HET-CAM) test system is used as a model of the cornea, conjunctiva, and iris to detect ocular corrosives and severe irritants. However, the CAM tissue structure is not similar to the cornea as the latter is not vascularized epithelium. Exposure of the rabbit eye to a chemical results in a pathophysiological reaction whereas the HET-CAM assay detects vascular injury. The differences in the structure of the CAM and the mammalian eye must be considered when using the HET-CAM assay as a predictor of potential for human eye irritation.

It is recognized that HET-CAM is an *in ovo* assay but for purposes of consistency, the term *in vitro* will be used when referring to this test method.

It is recommended that the draft HET-CAM BRD include discussions on:

- cellular mechanisms of corrosion and severe irritation (e.g., necrosis, apoptosis) and relevance to *in vitro* testing, and
- the role of responsive inflammatory cells in isolated rabbit eyes and how this compares to the responsive inflammatory cells in the CAM.

Furthermore, additional literature and laboratory research to review the following questions are recommended:

- How much and what kind of data are available for using eggs at incubation day 7?
- What is known about the development of the chorioallantoic membrane, its sensitivity and its reactivity on incubation day 7 compared to incubation day 9?
- What kinds of data about pain receptors are present on the CAM on either incubation day 7 or day 9?
- How does the incubation day affect the reliability and variability of the data?

1.1.2 Advantages and Limitations of Mechanisms/Modes of Action of the HET-CAM Test Method

The HET-CAM test method appears to be suitable as a limited screen for a broad array of different types of chemicals. A deficiency of the CAM is that it has no structures comparable to the iris and cornea. Chemical exposure in the rabbit eye can be relatively long (usually never washed) as compared to the HET-CAM assay, which is relatively short (5 minutes).

The actual endpoints assessed in the two test systems are different. The rabbit eye test assesses each specific major eye structure endpoints up to 21 days post exposure while the HET-CAM

test method uses a scoring system and formula to evaluate the degree of blood vessel hemorrhage, lysis, and coagulation.

1.1.3 Similarities and Differences of Mechanisms/Modes of Action and Target Tissues Between the HET-CAM Test Method and Humans and Rabbits

Much is known about differences in mechanisms/mode of action between the HET-CAM test method and humans and rabbits. All of these differences have to be considered and kept in mind as comparisons are made. Exposure of the rabbit conjunctiva to a chemical results in an immunological reaction whereas in the HET-CAM assay, the result is a measure of vessel necrosis. The differences in response of adult tissues (with a developed immune system) verses embryonic tissues (with a much undeveloped immune system) also need to be kept in mind when reviewing the results from the HET-CAM test method. Due to these differences, it cannot be assumed that adverse changes that occur in the HET-CAM test method are going to be similar to what may occur in the rabbit or human eye.

1.1.4 Mechanistic Similarities and Differences Between the HET-CAM Test Method, the *In Vivo* Rabbit Eye Test Method, and/or Human Chemically-Induced Eye Injuries

Due to the differences in the mechanisms of the response between the tests, the *in vivo* rabbit eye test will more closely predict what changes will occur in the human eye over a period of days. The *in vivo* rabbit test follows the eye over a period of up to 21 days and any long-term effects can be noted in endpoints very relevant to human exposure (iris, cornea, conjunctiva). Comparatively, the HET-CAM test method is a short-term test (5 minutes) with few endpoints (hemorrhage, lysis, coagulation) and no responses related to the cornea or iris.

Any relationship between the short-term effects observed in the HET-CAM test method to the long-term effects seen in rabbits or humans should be explored in the HET-CAM BRD. Such an evaluation may provide additional support for the use of the HET-CAM method to assess the delayed and long-term effects of corrosives and severe irritants.

1.2 Regulatory Rationale and Applicability

1.2.1 Similarities and Differences Between Endpoints Measured in the HET-CAM Test Method and the *In Vivo* Rabbit Eye Test Method

The endpoints are very different between the *in vivo* rabbit eye and the HET-CAM test methods. The *in vivo* rabbit eye endpoints are very similar, if not identical, to what may happen to a human eye after exposure to a substance. The HET-CAM endpoints are a representation of what may happen by inferring from the onset of blood vessel necrosis in the CAM.

1.2.2 Suggestions Regarding Other Evidence that Might be Used in a Tiered Testing Strategy

The BRD has summed up these issues as five criteria that must be achieved. Four of the five criteria seem to be achievable. One criterion, which may be difficult to achieve, is criterion number 4: “Provide improved prediction of adverse health effects in the human”. This criterion would be difficult to achieve unless comparative data are generated using substances from a standardized repository that are already known to cause specific effects in humans. The HET-

CAM assay and other identified assays must all be tested using the same standard substances to determine if the assay can improve the prediction of adverse eye effects for humans.

It is hard to visualize that the HET-CAM test method, in its current state of performance, would do more than add another level of testing which would rarely supplant the existing tests. Rather, the HET-CAM test method may have the potential to complement other tests in a tiered-testing approach.

2.0 HET-CAM TEST METHOD PROTOCOL COMPONENTS

2.1 Description and Rationale of the Components for the Recommended HET-CAM Test Method Protocol

The recommendations from the draft HET-CAM BRD appear to appropriately integrate protocol components and specific procedures from the various published literature. These BRD recommendations also include developing consistent scoring and calculation of irritation indices.

Reference substances that are part of the performance standards developed for the HET-CAM test method should be identified in the BRD. These reference substances would be used to evaluate test methods similar to HET-CAM. The HET-CAM BRD also should clarify the decision criteria for identifying ocular corrosives and severe irritants.

2.2 Basis for Selection of the HET-CAM Test Method System

Historically, the chick embryo has been extensively utilized. The specific strain, stock and age of White Leghorn eggs, which has been recommended in the BRD, is common and fairly easy to obtain; use of these eggs would provide consistency for the HET-CAM assay results.

2.3 Identification of Proprietary Components

The Panel agrees with the BRD, there are no proprietary components of the test system.

2.4 Numbers of Replicate and/or Repeat Experiments for Each Test

The BRD recommendations on the numbers of replicates and/or repeat experiments would provide uniformity and consistency to the HET-CAM assay in interpreting the results. Many alternative assays that are submitted to regulatory agencies have, as part of the protocol, a standardized number of replicates that must be used in order for the test system to be considered valid.

2.5 Study Acceptance Criteria for the HET-CAM Test Method

Since the study acceptance criteria varied between the various test method protocols, a definition of what constitutes a positive result is needed. Also, since there are times when the concurrent control can show quite a bit of variation, tabulation and use of historical control data need to be considered. More objective criteria for assessment would enhance the repeatability and

reliability of the HET-CAM test method. Objective criteria also would enhance the validity of interlaboratory comparisons.

2.6 Basis for Any Modifications made to the Original HET-CAM Test Method Protocol

The Panel agrees with the BRD recommendations on the bases for any modifications made to the original HET-CAM test method protocol.

2.7 Adequacy of the Recommended Standardized Protocol Components for the HET-CAM Test Method

The Panel agrees with the BRD recommendations for the development and use of a standardized HET-CAM test method protocol. A critical recommendation is the inclusion of BOTH concurrent negative and positive controls each time the assay is conducted. In addition, investigators need to accumulate historical data for their positive and negative controls in order to better define the range of positive and negative responses as different materials are tested in the HET-CAM assay.

3.0 SUBSTANCES USED FOR PREVIOUS VALIDATION STUDIES OF THE HET-CAM TEST METHOD

3.1 Substances/Products Used for Prior Validation Studies of the HET-CAM Test Method

The types and numbers of substances/products used in prior validation studies appear adequate.

3.2 Coding Procedures Used in the Validation Studies

It was difficult to determine if the coding procedures used in the validation studies were appropriate. There was not enough information to determine the appropriateness of the coding used. As long as the quality and multiplicity of sources of the data were sufficient to draw meaningful conclusions, it does not matter if coding was not used.

4.0 *IN VIVO* REFERENCE DATA USED FOR AN ASSESSMENT OF TEST METHOD ACCURACY

4.1 *In Vivo* Rabbit Eye Test Method Protocol(s) Used to Generate Reference Data

The *in vivo* rabbit eye test method protocol(s) used to generate reference data in the cited studies were appropriate.

4.2 Interpretation of the Results of the *In Vivo* Rabbit Eye Tests

The interpretation of the results of the *in vivo* rabbit eye tests was correct. The *in vivo* methods described have been judged by the agencies using these methods as suitable for their regulatory

needs. The concern can reasonably be raised that these regulatory classification methods may be less than adequate for use in evaluating or making distinctions between *in vitro* methods and their suitability for chemical or product class evaluations.

4.3 *In Vivo* Rabbit Eye Test Data Quality with Respect to Availability of Original Study Records

If there are a few test substances that lack original study records, then they should not be given the same weight as those test substances with original study records. However, if there are many test substances that lack original study records and it appears that obtaining the original study records may be difficult, then such studies should be given equal weight with those that have original study records. In the case of the HET-CAM test method, original study data (e.g., laboratory notebooks) were not available for any of the reports evaluated. However, a lack of original study records does not necessarily raise concerns about a study. As long as an evaluation of the results can be made and the quality of the study otherwise is adequate (as is the case for the studies evaluated in the HET-CAM BRD), the study should be used.

4.4 *In Vivo* Rabbit Eye Test Data Quality with Respect to GLP Compliance

The criteria used in selecting agents in some of the studies for the HET-CAM test method cited in the BRD were not specified. The Balls et al. (1995) project included the criterion that the *in vivo* data were from GLP-compliant post-1981 studies conducted in accordance with OECD TG 405 (OECD 1987). The Spielmann et al. (1996) project was conducted under blind conditions according to GLP standards in laboratories of the chemical and drug industry in Germany. The Panel recommends that the status or availability of additional information on GLP compliance needs to be pursued more diligently.

However, as the GLP regulations do not deal with the actual performance of the tests as much as with documentation, no distinction needs to be made in the weight given to GLP-compliant versus non-GLP-compliant studies in the BRD as long as the work was performed in well-established laboratories (e.g., stable workforce, significant throughput in that section of the laboratory, long term experience with the test method, historical data, adequate supervisory staff). It is recognized that these are some characteristics of a well-established laboratory and are not meant to be criteria for determining such laboratories. Furthermore, according to the current EU and OECD documents on the validation of toxicity tests, when the basic requirements of the GLP procedure (the "spirit" of GLPs) have been implemented in a study, lack of complete/formal GLP compliance is not an adequate criteria to exclude *in vivo* or *in vitro* data from the evaluation of the performance of a toxicity test.

4.5 Availability of Relevant Human Ocular Toxicity Information

The small set of human data, whether from accident reports or controlled human studies, is of little value in examining the performance of an *in vitro* test. Appropriately, the discussion of this topic is quite limited. Very little human ocular injury data have been accessed and most of the available information originates from accidental exposure for which the dose and exposure period were not clearly documented. Accidental exposures have no measure of dose and

typically, even if the individual is seen in a clinical setting, there is no "scoring" or time course data.

However, it would seem worthwhile to determine if the current ocular hazard classification schemes are working correctly to protect workers and the public from severe eye injury. While it is difficult to obtain specific data from the various databases, they can be useful to give reassurance that current schemes appear to be protecting the public. According to the European Cosmetics, Toiletries and Perfumery Association (COLIPA) Task Force on Eye Irritation workshop report (Bruner et al. 1998), the International Life Sciences Institute (ILSI) has published a human eye irritation classification scheme (see Table II in Bruner et al. 1998) and planned to search databases on human eye irritation. Therefore, it is recommended that COLIPA and ILSI be consulted for human data.

The Panel also recommends that a greater effort be made to obtain, consider, and use information on human topical ocular chemical injury. The USEIR may be one source of such information. Literature sources of human topical ocular chemical injury include, but are not limited to, Grant (1974), Fox and Boyes (2001), and Fraunfelder (1982).

4.6 Accuracy and Reliability of the *In Vivo* Rabbit Eye Test

There should be more discussion in the HET-CAM BRD of the variability of the rabbit data. This is particularly important in the determination of the accuracy of an *in vitro* test method. While there are often multiple results for each *in vitro* determination of irritation potential, there is generally only one *in vivo* test result. Because of the known variability in the rabbit test (e.g., Weil and Scala 1971; Spielmann 1996), it is not possible from the data presented to determine if the inconsistencies between the two tests are due to "failure" of the *in vitro* test method or a misclassification by the single *in vivo* result provided.

When interpreting the *in vitro* test data, the differences in reproducibility/variability of the *in vivo* Draize eye test data have to be taken into account. Therefore, before data analysis is performed, it has to be defined how this special feature of the Draize eye test will be taken into account when comparing it to results from *in vitro* tests and when attempting to determine the predictive value of the *in vitro* alternatives.

This important aspect has been cited as a reason why the replacement of the Draize eye test by *in vitro* tests has failed in the past. Although it is well documented in the scientific literature (e.g., Figure 1 in Balls et al. [1995]) and in a review by Spielmann (1997), additional discussion in the HET-CAM BRD is warranted.

The Draize eye irritation test has never gone through a validation process. However, data on the reliability of the *in vivo* rabbit eye test do exist in the literature, most notably the intra- and inter-laboratory study published by Weil and Scala (1971). Using a fixed protocol and a single supply of chemical agents tested in 25 laboratories, these investigators identified "good" laboratories as those, which had the lowest variance in ranking of irritancy using a sum of ranks statistical measure. They also found that nonirritants provided little useful information on laboratory

performance. GLP regulations were not in place at the time of this study, but are not thought to be critical in the evaluation of the data.

Using data from the Weil and Scala (1971) study, another evaluation showed the difference in MAS values that can be obtained between different laboratories. For three of the ten substances tested, the *in vivo* Draize eye irritation test indicated that the substances were classified as nonirritant (MAS < 20) to irritant (MAS > 60) when tested in 24 laboratories (Spielmann 1996).

It is documented that the Draize eye test has low variability at both ends of the MAS scale (e.g., the low end in the range of non-irritating chemicals and at the upper end of the scale in the range of severely eye irritating materials) (Spielmann 1996). However, in the middle range, the variability is very high for such substances (as indicated by the high CV and SD values in Balls et al. [1995]). While any repeat performance of *in vivo* rabbit eye irritancy testings or testing of known corrosives or severe irritants should be strongly discouraged, it is important to have available multiple *in vivo* rabbit eye test data that demonstrate reproducible results.

In the development of alternative methods to intact animal testing, the question always arises regarding the quality of reference *in vivo* data used to evaluate or validate the newer *in vitro* method. These questions typically center on two major concepts. The first is the availability of a reference standard for measuring the intended effect. The second is the reproducibility and reliability of the *in vivo* test. With respect to ocular injury (irritation or corrosion), there is no "gold standard". That is, there is no set of substances that have been shown, regularly and reproducibly, in any competent laboratory, to produce a particular degree of irritancy or damage in the intact rabbit eye. Consequently, the evaluation (or acceptability) of an alternative method is unavoidably biased by the selection of the *in vivo* data used in that evaluation.

Not all substances evaluated in the HET-CAM BRD were tested concurrently in both the *in vivo* rabbit eye and the HET-CAM test methods. In addition, none of the substances were identified as having been tested in the *in vivo* rabbit eye test in multiple laboratories. It would seem that the entire effort to develop alternatives to intact animal testing for ocular effects would benefit from some attention to providing an approximation of a "gold standard".

An effort should be made to determine if the *in vivo* results are consistent with the known toxicity of these materials (e.g., as indicated in the RTECS or IUCLID databases) would be useful. It is imperative that a greater effort be made to access suitable human data from other sources such as Hazardous Substances Data Bank (HSDB), the Physician's Desk Reference (PDR) and the Poison Control Center network.

The Panel recommends that any future optimization and validation studies should use existing animal data if they are available. If important data gaps are identified, additional animal studies should only be conducted with the minimum number of animals. Such studies should be carefully designed to maximize the amount of pathophysiological information obtained and conducted under GLP conditions.

Minority Opinion
This section was approved by consensus of the Panel with a minority opinion from Dr. Martin Stephens that sufficient animal data are available for further optimization/validation studies and no further animal testing should be conducted (see Minority Opinion from Dr. Stephens in **Section IV - 12.3**).

5.0 HET-CAM TEST METHOD DATA AND RESULTS

5.1 HET-CAM Test Method Protocols Used to Generate Data Considered in the BRD

The test method protocols used to generate each set of data considered in the BRD were adequately described. It is recommended that the type of irritation score (IS) (A or B) analysis method used by each study be detailed in Section 5.4 of the HET-CAM BRD.

5.2 Comparative HET-CAM Test Method–*In Vivo* Rabbit Eye Test Data Not Considered in the BRD

For the validation of the BCOP test method (Gautheron et al. 1994), an *in vivo* study was performed by one laboratory. Draize data from this *in vivo* study may be a source of data that could be used in the BRD evaluation for available HET-CAM data.

5.3 Statistical and Nonstatistical Approaches Used to Evaluate HET-CAM Data in the BRD

The approaches used to evaluate the HET-CAM data appear to adequately describe the accuracy and reliability of the test method. However, given the unavailability of original HET-CAM data, a definitive statement regarding the adequacy of these approaches is not feasible.

The accuracy analysis was complicated by a lack of consistent test and evaluation methods in the literature. Analysis methods in the HET-CAM BRD include the IS(A), IS(B), Q-Score, S-Score, and IS and Irritation Threshold Concentration scores, or in some cases, just classifications based on any of these analysis methods. Results were reformulated in the BRD to be consistent with regulatory agency classifications. The procedure was as good as possible given the lack of consistency among studies. This certainly is not optimal and more internally consistent data are needed.

The classification criteria using these analysis methods should be optimized, including considering the formula for combining information and the irritancy categorization of that result.

5.4 Use of Coded Substances, Blinded Studies, and Adherence to GLP Guidelines

Whether coded chemicals were tested, or the identity of the chemicals is unknown is adequately documented (HET-CAM BRD Section 3.4). Whether GLP guidelines were followed is detailed in Appendix B of the BRD. How well the studies followed GLP guidelines cannot be determined from the studies. In most of the studies, quality assurance was likely not involved. If

studies were conducted following GLP principles, which is likely the case for most of the studies, they should be accepted. GLP-criteria should not overrule all the other criteria for final acceptance of studies for retrospective validation of the HET-CAM test.

Ideally minimal criteria or requirements, such as (1) a well described materials and methods section and (2) criteria for a corrosive or severe irritant call, should be provided and be used to determine an adequate study. However, it is recognized that not all studies would provide such information. Consequently, as long as the data from the study can be interpreted and does not have any serious deficiencies, such as inadequate number of animals, it should be acceptable.

5.5 "Lot-to-Lot" Consistency of the Test Substances and Time Frame of the Various Studies

There is not enough information on "lot-to-lot" consistency. It is expected that different batches of substances may give some quantitative differences in irritation classification results but a major qualitative difference in irritation classification would not be expected (i.e., classification of a highly severe substance should remain severe between batches of substances). When the irritancy classification of a substance is on the borderline between nonsevere irritant and severe irritant, "lot-to-lot" variations may have an effect on the results. In other words, one batch of a borderline substance may produce a severe irritant response while another batch may produce a nonsevere irritant response.

6.0 HET-CAM TEST METHOD ACCURACY

6.1 Accuracy Evaluation of the HET-CAM Test Method for Identifying Ocular Corrosives and Severe Irritants

The accuracy of the *in vitro* test using the different evaluation criteria has been adequately evaluated. Accuracy evaluations were limited to the substances evaluated in nine *in vitro-in vivo* comparative studies.

1. Accuracy was assessed separately for each *in vitro-in vivo* comparative study.
2. Accuracy was assessed after pooling data across comparative studies that used the same method of data collection and analysis.

Overall, false positive rates ranged from 20% (8/40) to 27% (12/45) and false negative rates from 0% (0/12) to 7% (1/14) compared with *in vivo* rabbit eye test method data classified according to the GHS (UN 2003), the EPA (1996), or the EU (2001) ocular irritancy classification systems. To what degree false results can be reduced with more replicates, more understanding of the various sources of variability, and further optimization of the categorization decision rule is unclear. It will be essential to identify which structural classes of chemicals this test system works for and which ones it performs poorly for.

Tables 6-1 to 6-3 and Table 6-7 of the HET-CAM BRD are quite helpful in summarizing results on all the required accuracy measurements and give a good overview of the performance of the HET-CAM test method. HET-CAM BRD Table 6-9 provides clear information on discordant results, which also are well described in the text.

In addition to the analyses conducted in the BRD, the Panel recommends an assessment based on ranking of experimental data for severity for both the reference method and the *in vitro* test be conducted.

Minority Opinion
Drs. Martin Stephens and Peter Theran note that the term "accuracy" is used throughout the four BRDs and this Panel Report to address the degree of consistency between the *in vivo* rabbit (Draize) test and each of the four *in vitro* alternative test methods being evaluated.

It is well documented that there is a significant degree of variability in the data produced by the *in vivo* rabbit eye test when it is compared with itself, which raises the question as to the accuracy of the *in vivo* test to predict the human experience. Given this variability and the fact that no data demonstrating the ability of the *in vivo* test to predict the human experience was presented to the Panel, Drs. Stephens and Theran feel it should be recognized that this test is an imperfect standard against which the new tests are being measured.

Drs. Stephens and Theran are filing a minority report because they believe that the term "accuracy" is inappropriately used, and that it is more appropriate to use the term "consistency with *in vivo* data" when comparing test results.

6.2 Strengths and Limitations of the HET-CAM Test Method

Concordance assessments are severely limited by the lack of reported data and the differences between methods and analysis methods used in the different studies. False positives and false negatives are identified where possible. Categorization methods used by the authors in the original studies were not designed to meet regulatory agencies requirements. These limitations are clearly spelled out.

It is known that there is much variability among Draize data (Weil and Scala 1971; Spielmann 1996). In the case where an *in vitro* classification is different from the *in vivo* classification, the variability of the *in vivo* response should be reviewed.

6.3 HET-CAM Test Data Interpretation

Because of the limited nature of the reported data, considerable effort was necessary to interpret the data. Data interpretation and specific endpoints applied are sufficiently detailed, to the level possible. The description makes the reader quickly aware that the IS(B) analysis method is the best one to identify most ocular corrosives and severe irritants. A standardized test method is needed to produce more interpretable and consistent data.

It is recommended that IS(B) data of non-accepted studies (HET-CAM BRD Section 9.0) be compiled into a table to see what the outcomes are in these studies.

7.0 HET-CAM TEST METHOD RELIABILITY (REPEATABILITY/ REPRODUCIBILITY)

7.1 Selection Rationale for the Substances Used in the HET-CAM Test Method Reliability Assessment

The rationale for compound selection is based primarily on the easy availability of *in vivo* rabbit eye data. The quality of such data is a weakness for all *in vitro* validation studies. A rationale based on the quality of *in vivo* data (after a thorough investigation and independent checks) would have been better. Selection of substances of which *in vivo* irritancy grades are confirmed by at least two studies would have given more power to the validation of HET-CAM and other test methods. The Panel notes that the above limitations are limitations of the studies used in the analysis and thus limitations of the analysis in the HET-CAM BRD.

7.2 Intralaboratory Repeatability and Intra- and Inter-laboratory Reproducibility of the HET-CAM Test Method

Analysis on intralaboratory repeatability and intralaboratory reproducibility could not be done due to lack of available data at the time of BRD development. This is a weakness in the validation of the HET-CAM, but should not be a roadblock for its use in identifying ocular corrosives and severe irritants.

Qualitative and quantitative analysis on the interlaboratory variability was well addressed in the HET-CAM BRD. Interlaboratory data were available from four to five laboratories. Ocular irritancy classifications from HET-CAM studies are compared to *in vivo* rabbit eye classifications for each agency classification system. Comparisons are given in HET-CAM BRD Tables 7-1 to 7-3. The participating laboratories agreed on at least half the calls and total agreement occurred frequently. This analysis shows that less agreement among laboratories is obtained with nonsevere irritants/nonirritants. The interlaboratory correlations given in BRD Table 7-7 (for Balls et al. [1995]) vary considerably; S-Scores for chemicals insoluble in water range from -0.9 to 0.852. Clearly, additional work is needed to improve interlaboratory consistency, when using the S-Score analysis method.

Use of %CV values has limitations when evaluating a narrow range of scores (i.e., 0-21 for the HET-CAM test method). Alternative approaches for measuring reproducibility (intra- and inter-laboratory) could be used and are recommended. One approach to assess variability could be the use of a non-parametric analysis, which is useful for small sample sizes and when the data may well not be normally distributed. The Kruskal-Wallis and Mann-Whitney tests evaluate for differences between groups, K groups (where $K > 2$ groups) and 2 groups, respectively. These tests are appropriate for comparing data with continuous outcomes, such as the IS score, to answer the question "do scores differ between laboratories" when comparing replicate scores from the same substance. An assumption for both tests is that observations are independent and identically distributed, and this would not be the case for different substances. So these tests would be useful for a substance-by-substance evaluation if the raw data are, or can be made, available.

A chi-square test for homogeneity of substances across laboratories may be used. But there are so many test substances that this test will not perform well. One could test whether the proportions of substances called severe significantly differ between the laboratories. For HET-CAM, there are enough substances to assume normality of the proportions, so one could do a global test for differences and then use one of a variety of methods for assessing multiple comparisons if the global test for no difference is rejected. This would be a straightforward measure of laboratory differences.

The Spearman rank correlation also is a good non-parametric measure of correlation. It would apply to the IS scores, but not to severe versus not severe outcomes.

The following items are noted for revision in the HET-CAM BRD:

- In BRD Tables 7-4 and 7-5, it would be helpful to have the sample size noted in the table to verify understanding of the text (this is true for some other tables as well). There is nothing in the Table heading or footnotes that say measurements are taken across laboratories.
- Motivation for inclusion of Balls et al (1995) was given. This should also be done for Hagino et al. (1999) on BRD page 7-2 (line 36).
- BRD Page 7-16, line 339: reference is made to Ohno et al. (1999) but no information on this publication can be found in Appendix B.

7.3 Availability of Historical Control Data

The absence of historical negative and positive control data is a weakness in the validation of the HET-CAM test method but this should not be a roadblock for the acceptance of this model as alternative test to detect ocular corrosives and severe irritants.

The Panel notes that some non-accepted publications (HET-CAM BRD Section 9) included positive controls. These publications may give some more information on the reproducibility of HET-CAM. Gilleron et al. (1996, 1997) included a positive control in all HET-CAM studies. Historical control data (90 tests with 0.9% NaCl as negative control, 80 tests with *N,N*-dimethylformamide as a negative control, and 15 tests with imidazole as a positive control) were obtained from Johnson & Johnson Pharmaceutical Research and Development laboratories (Beerse, Belgium) to assess intralaboratory reproducibility. The fact that a test substance applicator was used (which is different from all the other studies discussed in the BRD) should not influence the outcome of the study.

It also is noted that some studies used positive controls that are typically considered nonirritants. Appropriate recommendations are made for the use of concurrent positive and negative controls in the HET-CAM BRD.

7.4 Effect of Minor Protocol Changes on Transferability of the HET-CAM Test Method

The sensitivity of the method to minor protocol changes is impossible to evaluate without having more standardized studies with measures of variability.

Optimization and validation studies are needed for routine regulatory use for hazard classification. Accuracy and reliability may be improved by tailoring the *in vitro* classification scheme to the classification systems of the regulatory agencies and further optimizing the criteria for these systems.

8.0 TEST METHOD DATA QUALITY

8.1 Impact of GLP Noncompliance and Lack of Coded Chemical Use

As scoring of the effects is still somewhat subjective, knowledge of the substances might have influenced scoring of the endpoints during the conduct of the *in vitro* test. Failure to use GLP guidelines may have had a qualitative impact on borderline classifications of nonsevere/severe irritants. The use of GLP guidelines assures that there was good control of the test system, acceptance criteria were defined, evaluation criteria were defined, and there were data audits. Lack of GLP compliance may be overcome by use of coded substances and appropriate data handling.

The Panel recommends that information on coding provided in Section 3.4 of the HET-CAM BRD also be included in Appendix B2.

8.2 Results of Data Quality Audits

The Panel agrees that no data quality checks could be done. This is a weakness not only for the HET-CAM validation but probably also for all other tests as a data quality check is included in the GLP guidelines where an independent group (Quality Assurance Unit; QAU) performs this task. Involvement of QAU is rarely included in validation studies.

8.3 Impact of GLP Deviations in the Data Quality Audits

As this cannot be deduced from the available information, the Panel agrees with the BRD conclusion that the impact of the deviations from GLP guidelines cannot be evaluated.

8.4 Availability of Original Records for an Independent Audit

The Panel agrees that the availability of laboratory notebooks or other records is adequately discussed in the BRD. Evaluation presented in the BRD has been done with the available data and information. The ICCVAM recommendation that all of the data supporting validation of a test method be available with the detailed protocol under which the data were produced is reasonable and should be supported (ICCVAM 2003). Availability of notebooks or other records would increase the "trust index" of the conclusions presented in the HET-CAM BRD.

9.0 OTHER SCIENTIFIC REPORTS AND REVIEWS

9.1 Other Published or Unpublished Studies Conducted Using the HET-CAM Test Method

The Panel agrees that a comprehensive review is made on available publications. The Panel wonders if the criteria for acceptance of literature for evaluation were too strict and relaxing the criteria would have allowed more studies to be included in the final evaluation discussed in the BRD. Additionally, by requesting some additional information on publications closely satisfying the inclusion criteria might have resulted in more studies considered for final evaluation of the performance of the HET-CAM test.

It is recommended that an evaluation on the impact of relaxing the data inclusion criteria be conducted, and additional resources should be placed on contacting authors of relevant papers and individuals that may have *in vitro* and/or *in vivo* data that may be used in the evaluation of the performance of HET-CAM. Additionally, it is recommended that information be placed into the HET-CAM BRD that indicates from which publications additional information was obtained and from which publications additional information was not obtained.

9.2 Conclusions Published in Independent Peer-Reviewed Reports or Other Independent Scientific Reviews

The conclusions published in independent peer-reviewed reports and other independent scientific reviews were adequate and complete. It was useful to have the motivation for exclusion of the studies for the final evaluation on the performance of the HET-CAM test. But, once again, the criteria may have been too strict for inclusion of some studies.

Recommendations made by the Panel in **Section IV - 9.1** of this report are applicable to this section.

9.3 Approaches to Expedite the Acquisition of Additional Data

An approach to expedite the process for obtaining additional in-house data could be to make a review on *in vivo* data of a preferred list of compounds and ask companies if they can deliver additional data supporting or contradicting the conclusions made by the Panel.

10.0 ANIMAL WELFARE CONSIDERATIONS (REFINEMENT, REDUCTION, AND REPLACEMENT)

10.1 Extent to Which the HET-CAM Test Method Refines, Reduces, or Replaces Animal Use

This section of the HET-CAM BRD addresses many of the considerations relevant to the 3Rs of refinement, reduction, and replacement. However, the discussion of some issues seems incomplete. In addition, other animal welfare considerations (perhaps not explicitly related to the 3Rs) still need to be discussed, or at least mentioned.

- It is recognized that HET-CAM is an *in ovo* assay but for purposes of consistency the term *in vitro* will be used when referring to this test method.
- Section 10.0 of the HET-CAM BRD mentions that pain perception is unlikely to occur prior to incubation day 9. It is recommended that discussion on pain perception (as is discussed in Section 2 of the BRD) in this section should be expanded.
- It is recommended that Section 10.0 of the HET-CAM BRD also mention the tiered-testing strategy that is being envisioned, namely, the use of HET-CAM test as a first tier test and *in vivo* testing as the second tier, triggered only by a negative finding in the first tier. Thus animals would be needed only to confirm the absence of a severe or corrosive response in the initial tier.
- Given HET-CAM's place in this potential two-tiered battery, the test method would probably best be considered a "partial replacement" in 3Rs parlance, albeit one that also results in refinement and reduction.
- In this section of the HET-CAM BRD or elsewhere, it should be stated that:
  - additional optimization and validation studies should rely on existing *in vivo* data
  - the low rate of false negatives (underpredictions) for HET-CAM has the animal welfare advantage of reducing the exposure of rabbits in the follow-on testing to severe irritants or corrosives
  - any test method optimization should seek to further decrease the false negative rate

11.0 PRACTICAL CONSIDERATIONS

11.1 HET-CAM Test Method Transferability

The proposed test method, as detailed in Appendix A of the HET-CAM BRD, should be readily transferable to properly equipped and staffed laboratories. A video on the method and on scoring would make implementation easier and ensure correct conduct of the test method.

11.1.1 Facilities and Major Fixed Equipment Needed to Conduct the HET-CAM Test Method

The Panel agrees with the BRD on the facilities and major fixed equipment needed to conduct the HET-CAM test method. All the equipment and supplies seem to be readily available. In addition, technicians who are trained in the assay do not need to be trained in proper animal handling techniques, husbandry and all the other regulatory issues that arise when intact animals need to be housed and used.

11.1.2 General Availability of Other Necessary Equipment and Supplies

The Panel agrees with the BRD on the general availability of other necessary equipment and supplies.

11.2 HET-CAM Test Method Training

11.2.1 Required Training to Conduct the HET-CAM Test Method

The Panel agrees with the BRD on the required level of training and expertise needed for personnel to conduct the HET-CAM test method. In addition, training on the HET-CAM assay should involve both positive and negative controls, identifying the critical endpoints, and calculating the irritation indices.

11.2.2 Training Requirements Needed to Demonstrate Proficiency

The Panel agrees with the BRD on the training requirements needed for personnel to demonstrate proficiency. In addition, some kind of limited refresher training should be conducted periodical (e.g., every 2 years). A training video and other visual media on the technical aspects of the assay is recommended. Training approaches in the application of this test method should be developed and implemented for use in training.

11.3 Relative Cost of the HET-CAM Test Method

The Panel agrees with the BRD on the costs involved in conducting the *in vivo* test. Rabbit test costs are consistent with past experience.

11.4 Relative Time Needed to Conduct a Study Using the HET-CAM Test Method

The Panel agrees with the BRD on the amount of time needed to conduct a study. The duration of the *in vivo* rabbit eye test is consistent with past experience. However, it is recognized that a corrosive or severe irritant may be detected within a few hours using a single rabbit.

12.0 PROPOSED TEST METHOD RECOMMENDATIONS

12.1 Recommended Version of the HET-CAM Test Method

12.1.1 Most Appropriate Version of the HET-CAM Test Method for Use in a Tiered Testing Strategy to Detect Ocular Corrosives and Severe Irritants and/or for Optimization and Validation Studies

The Panel agrees that the most appropriate version of the HET-CAM test method for use in a tiered-testing strategy and/or optimization and validation studies is the test method protocol recommended in the HET-CAM BRD. It is recommended that for the purpose of detecting severe eye irritants in the tiered-testing scheme outlined in the BRD, the HET-CAM test is useful for identification of severe or corrosive ocular irritants with the caveat that the HET-CAM has a high false positive rate. Positive results could be re-tested in a modified HET-CAM test method (e.g. using a lower concentration of test substance) to confirm the results. Alternatively, the positive substance could be tested in a different *in vitro* test method (e.g., ICE, IRE, BCOP). It is noted that data and information on the use of lower concentrations of test substances in the HET-CAM test method exist. Such information should be included in the BRD.

The proposed HET-CAM standardized test method protocol is adapted from the one by Spielmann and Liebsch (INVITTOX 1992). The method contains a negative control, a solvent

control (if appropriate), a positive control and benchmark controls (if appropriate). Overall, the method is similar to those used by most investigators, but recommends using the time required for an endpoint to develop as the criteria for assessing irritation potential (Kalweit et al. 1987, 1990). The IS(B) method exhibited the highest accuracy rate (78%) and the lowest false negative rate (0%) in identifying ocular corrosives and severe irritants.

More specifically, the use of a standardized protocol in future studies will allow for new data to be generated, which will allow further evaluation of the usefulness and limitations of the recommended test method protocol. The proposed standardized HET-CAM test method protocol includes the use of concurrent positive and negative control test substances, whereas the published protocols are inconsistent on the use of such control test substances. Including concurrent control substances in the HET-CAM test method protocol allows for an assessment of experimental variability across time, establishment of a historical control database, and development of acceptance criteria for each test based on the positive control substance inducing an appropriate response. The test method protocol also recommends the inclusion of appropriate benchmark substances, where possible, to aid in evaluating the ocular irritancy potential of test substances of a specific chemical class, or for evaluating the relative irritancy potential of a test substance within a specific range of irritant responses.

When using this method for substance classification, substances producing positive results (e.g., HET-CAM score defined as corrosive or severe irritant) obtained from this test method can be used to classify a substance as an ocular corrosive or severe irritant. Substances producing negative results (e.g., HET-CAM score defined as nonirritant, mild irritant, or moderate irritant) obtained from this test method would follow the tiered testing strategy.

12.2 Recommended Standardized HET-CAM Test Method Protocol

12.2.1 Appropriateness of the Recommended Standardized HET-CAM Test Method Protocol and Suggested Modifications to Improve Performance

The Panel recommends that procedures for applying and removing solids from the CAM be included in the standardized test method protocol. Solid substances may adhere to the CAM and demolish the CAM upon removal. Therefore, procedures for evaluating solids in this test method should be included in the test method protocol provided in Appendix A of the HET-CAM BRD.

Further optimization of the recommended standardized test method protocol should be possible. Optimization should increase the accuracy of the HET-CAM test method by reducing the moderate false positive rate while maintaining the low false negative rate. Optimization also should increase the reliability of the HET-CAM test method. Therefore, a retrospective analysis should be conducted to determine if different decision criteria might enhance the accuracy and/or reliability of the test method for the detection of ocular corrosives and severe irritants, as defined by the EU (2001), GHS (UN 2003), and EPA (1996) classification systems. Since it appears that the appropriate data are not available, a subset of substances in the recommended list of reference substances (HET-CAM BRD Section 12.4) should be tested to provide the necessary data.

12.2.2 Other Endpoints that Should be Incorporated into the HET-CAM Test Method

Other endpoints may be considered for use with the HET-CAM test method, but inclusion of these endpoints should not block retrospective validation of the HET-CAM test method with the parameters previously used to evaluate eye irritation potential.

The endpoints evaluated in HET-CAM are quite different from those evaluated in ICE, IRE, and BCOP, the organotypic test methods. For example, all three organotypic test methods include an evaluation of corneal opacity. Comparatively, the endpoints used in HET-CAM (development of lysis, hemorrhages, and coagulation) are unique to this test method; their use is based on proposed physiological similarities between the CAM and various structures of the eye (i.e., conjunctiva, cornea). Further optimization of the HET-CAM test method for the detection of ocular corrosives and severe irritants may be possible by considering different endpoints (e.g., trypan blue absorption, antibody staining, membrane changes) for evaluation and inclusion in the calculation of irritancy potential. Some of these may be comparable to those of the IRE, ICE and BCOP methods: membrane swelling, dye retention, visual evaluation, and microscopic evaluation. These additional tests may help reduce the number of false positives with the HET-CAM test.

12.3 Recommended Optimization and Validation Studies

It is recommended that an evaluation to determine the relationship or predictability between the short-term effects observed in the HET-CAM and long-term effects observed in rabbits or humans be conducted. Such an evaluation may provide additional support for the use of the HET-CAM method to assess the delayed and long-term effects of corrosives and severe irritants.

12.3.1 Recommended Optimization Studies to Improve Performance of the Recommended HET-CAM Test Method Protocol

No optimization studies are needed to lower the false negative rate of the HET-CAM test method. However, studies to lower the false positive rate are needed. Optimization studies should make maximum use of retrospective analyses to preclude the need for further, time-consuming studies. Any further optimization and/or validation work should take full advantage of the modular approach to validation that the ECVAM is developing. The work could identify needed modules (e.g., interlaboratory reliability) and focus on gathering data for those needed modules. This would avoid the time and expense of a full-blown validation study.

It is recommended that any future optimization and validation studies should use existing animal data, if they are available. If important data gaps are identified, additional animal studies should only be conducted with the minimum number of animals. Such studies should be carefully designed to maximize the amount of pathophysiological (e.g., depth of injury) information obtained and conducted under GLP conditions. Any optimization and/or validation studies also should aim to minimize the number of animals used.

Optimization studies could increase the accuracy of the HET-CAM test method by reducing the moderate false positive rate while maintaining the low false negative rate. Therefore, a retrospective analysis should be conducted to determine if different decision criteria might enhance the accuracy of the test method for the detection of ocular corrosives and severe

irritants, as defined by the EU (2001), GHS (UN 2003), and EPA (1996) classification systems. Optimization studies also may involve the development of a protocol that includes re-testing of positive substances using a modified HET-CAM test method protocol, as described above.

It is noted that optimizing a method involves validation of the method only if the modifications do not have a major impact on the conduct of the study. The recommendation to optimize and to use an optimized method should not minimize the value of data already obtained with the method of Spielmann and Liebsch (INVITTOX 1992). As some laboratories already apply this method, the data generated in these laboratories should still be valid and be used for labeling of corrosives and severe irritants.

An optimized test method may be used when a positive finding is obtained in the HET-CAM test method of Spielmann and Liebsch (INVITTOX 1992); the optimized protocol should be applied as a second step. This optimized protocol should then be validated.

The high variability of the Draize test does not allow for 100% accuracy with any of the recommended optimized methods or any other proposal for change. Because not enough human data are available, reference is made to the Draize test. However, this test cannot be seen as a "gold standard" (see **Section IV - 4.6** of this report) and should be defined as a "reference standard".

The Panel also recommends that this BRD section should discuss the pros and cons of the immediate implementation of the HET-CAM test for ocular corrosion and severe irritation. For example, the discussion should answer the question: What, if anything, is the downside of foregoing the proposed optimization and validation work?

12.3.2 Recommended Validation Studies to Evaluate Performance of the Optimized HET-CAM Test Method Protocol

If optimization of the method is done to reduce the false positive rate and modifications have a major impact on the conduct of the study, a validation study should be done with the optimized method. As the false negative rate is 0%, it is recommended that validation of the optimized method to reduce the false positive rate while maintaining the low false negative rate.[3]

The Panel also recommends identification of reference substances that would be included as part of the performance standards developed for the HET-CAM test method. These reference substances would be used to evaluate optimized test methods that are similar to the HET-CAM test method.

[3] Practical use of the IS(B) method in pharmaceutical industry for other purposes: In the pharmaceutical industry, the IS(B) analysis method is used to assess irritating potential of nasal or intravenous formulations. In this respect the IS(B) analysis method was found to be very powerful to select the right formulations. Formulations that were identified as nonirritants by the IS(B) analysis method did not induce irritation in animals. Intravenous formulations, which came out as severe irritating in the IS(B) analysis method induced severe irritation in the blood veins of animals even with necrosis of blood vessel cells (Vanparys P, personal communications).

Minority Opinion

According to Dr. Martin Stephens, **Section IV – 12.3** recommends that additional optimization and/or validation studies be conducted, and the report leaves open the possibility of additional animal studies as part of this process. Dr. Stephens believes that no additional animal studies should be conducted for such optimization or validation exercises. He cited several reasons for holding this view:

1. Draize testing of severely irritating or corrosive chemicals causes extremely high levels of animal suffering.
2. The intended purpose of the alternatives under review is narrow in scope (i.e., simply to serve as a positive screen for severely irritating or corrosive chemicals). Negative chemicals go on to be tested in animals.
3. The Panel learned that more animal and alternative data exist that are relevant to each of the alternative methods, and greater efforts should be made to procure these and any other existing data.
4. Some relevant animal data were dismissed from the analysis of each alternative method, and this dismissal should be reevaluated in light of any need for additional data.
5. Suggestions for further optimization and/or validation studies should be assessed critically, in light of the fact that only the most promising alternative method need be developed further, not necessarily all four methods, and that whatever alternative is selected for further development need be optimized only to the point at which it is at least as good as the Draize test.
6. A new modular approach to validation has been developed that could potentially reduce the number of chemicals needed to fulfill each module. Such an approach, if pursued, might be workable with the data already summarized in the BRDs.

12.4 Proposed Reference Substances for Validation Studies

See **Section V**.

13.0 HET-CAM BRD REFERENCES

13.1 Relevant Publications Referenced in the BRD and Any Additional References that Should Be Included

It is recommended that the references in the public comments provided by Dr. med. Horst Spielmann, which lists relevant publications, should be included in the BRD.

14.0 PANEL REPORT REFERENCES

Balls M, Botham PA, Bruner LH, Spielmann H. 1995. The EC/HO international validation study on alternatives to the Draize eye irritation test. Toxicol In Vitro 9:871-929.

Bruner LH, de Silva O, Earl LK, Easty DL, Pape W, Spielmann H. 1998. Report on the COLIPA workshop on mechanisms of eye irritation. ATLA 26:811-820.

EPA. 1996. Label Review Manual. 2nd Edition. EPA737-B-96-001. Washington, DC:U.S. Environmental Protection Agency.

EU. 2001. Commission Directive 2001/59/EC of 6 August 2001 adapting to technical progress for the 28th time Council Directive 67/548/EEC on the approximation of the laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances. Official Journal of the European Communities L255:1-333.

Fox DA, Boyes WK. 2001. Toxic responses of the ocular and visual system. In:Casarett & Doull's Toxicology: The Basic Science of Poisons, 6th Edition. (Klaassen CD ed). New York:McGraw-Hill Press, 565-596.

Fraunfelder FT, ed. 1982. Drug-induced ocular side effects and drug interactions. Philadephia, PA:Lea & Febiger.

Gautheron P, Giroux J, Cottin M, Audegond L, Morilla A, Mayordomo-Blanco L, Tortajada A, Haynes G, Vericat JA, Pirovano R, Tos EG, Hagemann C, Vanparys P, Deknudt G, Jacobs G, Prinsen M, Kalweit S, Spielmann H. 1994. Interlaboratory assessment of the bovine corneal opacity and permeability (BCOP) assay. Toxicol In Vitro 8:381-392.

Gilleron L, Coecke S, Sysmans M, Hansen E, van Oproy S, Marzin D, van Cauteren H, Vanparys P. 1996. Evaluation of a modified HET-CAM assay as a screening test for eye irritancy. Toxicol In Vitro 10:431-446.

Gilleron L, Coecke S, Sysmans M, Hansen E, van Oproy S, Marzin D, van Cauteren H, Vanparys P. 1997. Evaluation of the HET-CAM-TSA method as an alternative to the Draize eye irritation test. Toxicol In Vitro 11:641-644.

Grant WM. 1974. Toxicology of the eye; drugs, chemicals, plants, venoms. Springfield, IL:Thomas.

Hagino S, Kinoshita S, Tani N, Nakamura T, Ono N, Konishi K, Iimura H, Kojima H, Ohno Y. 1999. Interlaboratory validation of in vitro eye irritation tests for cosmetic ingredients. (2) Chorioallantoic membrane (CAM) test. Toxicol In Vitro 13:99-113.

ICCVAM. 2003. ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods. NIH Publication No. 03-4508. Research Triangle Park, NC:National Institute of Environmental Health Sciences.

INVITTOX 1992. HET-CAM Test. ECVAM. Available: https://ecvam-sis.jrc.it/invittox/published/indexed_47.html [accessed 18 February 2004].

Kalweit S, Gerner I, Spielmann H. 1987. Validation project of alternatives for the Draize eye test. Mol Toxicol 1:597-603.

Kalweit S, Besoke R, Gerner I, Spielmann H. 1990. A national validation project of alternative methods to the Draize rabbit eye test. Toxicol In Vitro 4:702-706.

OECD. 1987. Acute Eye Irritation/Corrosion. Test Guideline 405. Paris, France: Organisation for Economic Co-operation and Development.

Ohno, Y, Kaneko T, Inoue T, Morikawa K, Yoshida T, Fuji A, Masuda M, Ohno T, Hayashi M, Momma J, Uchiyama T, Chiba K, Ikeda N, Imanashi Y, Itagaki H. 1999. Interlaboratory validation of the *in vitro* eye irritation tests for cosmetic ingredients. (1) Overview of the validation study and Draize scores for the evaluation of the tests. Toxicol In Vitro 13:73-98.

Spielmann H, Liebsch M, Kalweit S, Moldenhauer F, Wirnsberger T, Holzhutter H, Schneider B, Glaser S, Gerner I, Pape WJW, Kreiling R, Krauser K, Miltenburger HG, Steiling W, Luepke NP Müller N, Kreuzer H, Mürmann P, Spengler J, Bertram-Neis, E, Siegemund B, Wiebel F. 1996. Results of a validation study in Germany on two *in vitro* alternatives to the Draize eye irritation test, HET-CAM test and the 3T3 NRU cytotoxicity test. ATLA 24:741-858.

Spielmann H. 1996. Alternativen in der Toxikologie. In: Alternativen zu Tierexperimenten, Wissenschaftliche Herausforderung und Perspektiven (in German). (Gruber FP, Spielmann H, eds). Berlin/Heidelberg/Oxford:Spektrum Akademischer Verlag, 1006:108-126.

Spielmann H. 1997. Ocular Irritation. In: *In Vitro* Methods in Pharmaceutical Research. (Castell JV, Gómez-Lechón MJ, eds). London:Academic Press, 265–287.

UN. 2003. Globally Harmonised System of Classification and Labelling of Chemicals (GHS). New York & Geneva: United Nations.

Weil CS, Scala RA. 1971. Study of intra- and inter-laboratory variability in the results of rabbit eye and skin irritation tests. Toxicol Appl Pharmacol 19:276-360.